

Entwurf eines geographisch-historischen Informationssystems: GEOHIST

Pierau, Karl

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Pierau, K. (1993). Entwurf eines geographisch-historischen Informationssystems: GEOHIST. *Historical Social Research*, 18(4), 49-75. <https://doi.org/10.12759/hsr.18.1993.4.49-75>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Entwurf eines geographisch-historischen Informationssystems: GEOHIST

*Karl Pierau**

Abstract: Semantic data models are the connecting link between users' view of data and the physical realisation in the data base. The Entity-Relationship model is the best known type of semantic data models. The method is demonstrated on the basis of the development of a geographic-historical Information System.

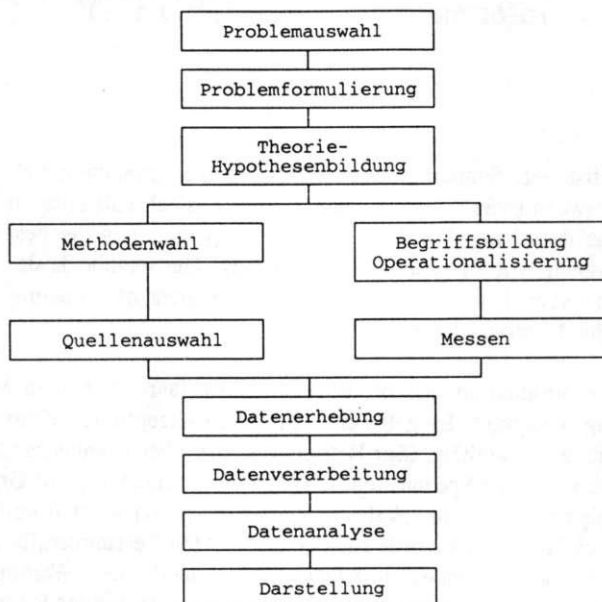
Die Software-Produktion war bis Ende der 60er Jahre in hohem Maße von Einzellösungen geprägt. Es gab keine allgemein akzeptierten Konzepte über Strategien ihrer Herstellung, über Betriebssysteme, über gewünschte Software-Eigenschaften. Wenige Spezialisten schrieben Programme, die auf Grund einer engen Problemsicht oft nur lokal und temporär genutzt werden konnten. Die Qualität eines Programms wurde nach beanspruchten Ressourcen für Speicherbedarf und Rechenzeit gemessen. Kriterien wie Änderbarkeit, Wartungsfreundlichkeit, die heute angelegt werden, spielten damals eine geringe Rolle. Es hing von der 'Kunst' des Programmierers ab, seinen Code so zu gestalten, daß er genial kurz, den beschränkten Ressourcen angepaßt, für andere aber vollkommen verständlich war.

In den folgenden Jahren wurde eine Reihe von methodischen Ansätzen zu den verschiedensten Bereichen der Software-Herstellung entwickelt. Ein sehr wesentlicher Schritt war, daß man eine Folge von wohlunterschiedenen Etappen bei der Herstellung von Software erkannte, die nacheinander abgearbeitet und parallel dazu dokumentiert werden sollten. Diese Etappen sind unter dem Begriff Phasenkonzept in die Literatur eingegangen. Für jede Phase wurden Methodiken und Arbeitsschritte entworfen.

Ähnliche Konzepte findet man in der Wissenschaftsmethodik. Speziell in der empirischen Sozialforschung wurden Forschungsstrategien entwickelt, die viele Analogien zum Phasenkonzept der Softwareherstellung aufweisen (vgl. Abb. 1). Verwiesen sei hier exemplarisch auf HSR, Supplement 1.

* Address all communications to Karl Pierau, Zentralarchiv für Empirische Sozialforschung, Abt. Zentrum für Historische Sozialforschung, Universität zu Köln, Lilien-cronstr. 6, D-50931 Köln.

Abb. 1 Forschungsprozeß der Historischen Sozialforschung



Datenbankentwurfsprozeß

Große Datenmengen zu verwalten, sie zu pflegen und auf sie zuzugreifen ist kein triviales Problem. Von führenden Informatikern wurde es als das Jahrhundertproblem der Informatik bezeichnet. Schon in den sechziger Jahren entstanden die ersten Programmsysteme (Datenbankmanagementsysteme - DBMS), die den Nutzern das Speichern, Pflegen und Wiederfinden von großen Datenmengen erleichtern sollten. Eines der ersten war das 1968 entwickelte hierarchische Datenbanksystem IMS von IBM. Anfang der 70er Jahre definierte das CODASYL- Komitee¹ eine Norm für Netzwerkdatenbanken [1].

Die nach dieser Norm konzipierten Systeme waren wesentlich flexibler als die hierarchischen. Im Bericht wurde empfohlen, zwei getrennte Sprachen für Datenmanipulation und für Datendefinition zu entwickeln. Die CODASYL-Norm definierte die Syntax der Datendefinitionssprache (DDL - Data Definition Language) und der Datenmanipulationssprache (DML - Data Manipulation Language). Diese netzwerkorientierten und hierarchischen Datenbanken

¹ Conference on Data Systems Language

benutzen aufwendige Datenbankadressen, die Änderungen im Datenbestand relativ kompliziert machen.

E.F. Codd entwickelte das Konzept der relationalen Datenbanken [2],[3]. Er formulierte erst 12 Regeln, die ein System erfüllen muß, um sich relational nennen zu können. Nach seiner Veröffentlichung von 1990 sind es inzwischen 333 Regeln. Das relationale Modell von Codd definierte nicht eine Sprache, sondern eine mathematisch fundierte Algebra zur Verknüpfung von Mengen.

Anfang der 70er Jahre wurde vom Nationalen Institut für Standardisierung der USA (ANSI) eine Gruppe von führenden Informatikern beauftragt, Standards für relationale DBMS zu entwickeln. Sie beschäftigten sich einige Jahre mit diesem Problem und stellten fest, daß man keine Standards für relationale DBMS vorgeben sollte, sondern nur für den Entwurf von Datenbanken. Sie schlugen vor, von drei verschiedenen Ebenen aus eine Sicht auf die Daten zu gewinnen [4]:

- externe Ebenen (Benutzersicht)
- konzeptuelle Ebene (logische Gesamtsicht)
- interne Ebene (Physische Sicht)

Auf jeder dieser Ebenen wird eine formalisierte Darstellung geliefert, die als Schema bezeichnet wird.

Welchen Vorteil hat eine solche Dreiteilung? Das konzeptuelle Schema ist ein Modell des Forschungsgegenstandes. Es ist vollkommen unabhängig von spezieller Hard- oder Software. Ausgehend von der Aufgabenstellung wird mit systemtheoretischen Mitteln ein Abbild von der zu untersuchenden Miniwelt geschaffen. Bei geeigneter formaler Darstellung des konzeptuellen Modells ist es möglich, formalen Regeln folgend, dieses Modell in konkrete Datenstrukturen, relational oder netzwerkorientiert, zu übertragen. Sich ändernde Hard- oder Softwarebedingungen, beispielsweise die Übertragung des Datenbestandes in ein anderes Datenbankmanagementsystem, lassen sich relativ einfach bewerkstelligen. Man geht erneut den formalisierten Weg vom konzeptuellen Schema zum internen Schema, ohne die ersten beiden Schemata ändern zu müssen. Ändert sich andererseits die externe Ebene und damit das externe Schema durch veränderte Aufgabenstellung oder durch eine veränderte 'Miniwelt', muß i.a. nicht das gesamte konzeptuelle Modell neu entwickelt werden, sondern nur bestimmte, von der Änderung betroffene, Bereiche.

Diese Dreiteilung trägt der Erkenntnis Rechnung, daß jedes Softwareprodukt nach der Fertigstellung einen Entwicklungsprozeß, der durch Änderung und Anpassung geprägt ist, durchläuft. Die Kosten von Software werden in wesentlich stärkerem Maße durch Pflege (Anpassung, Weiterentwicklung, Fehlerbeseitigung während der Nutzungsphase) beeinflusst, als durch Entwicklungsaufwand. Daher sind Übersichtlichkeit, Nachvollziehbarkeit des Entwurfsprozesses dringende Forderungen. Kein genialer Programmcode entscheidet über die Güte der Software, sondern ein nachvollziehbarer Entwurf.

Die Notwendigkeit eines nachvollziehbaren Entwurfs, das 3-Ebenenmodell und andere methodische Aspekte führten zur Begründung von Arbeitsschritten bei der Softwareherstellung, dem bereits erwähnte Phasenkonzept, das für eine Klasse von ähnlichen Ansätzen steht. Es kann wie folgt skizziert werden:

1. Ermittlung und Analyse des Informationsbedarfs

Welche Informationen der 'Umwelt' sind für die Anwendung von Interesse (eine 'Miniwelt' ist festzulegen). Der Inhalt des Informationssystems ist festzulegen.

2. Beschreibung der ermittelten Informationen

Die im ersten Schritt ermittelten Informationen liegen in informaler Form vor. Das kann zu Widersprüchlichkeiten (Inkonsistenzen) und Unvollständigkeiten führen. Man sollte daher Beschreibungsmittel benutzen, die die Ergebnisse aus dem erstem Schritt einheitlich und formalisiert, unabhängig vom DBMS, darstellen. Diese semantische Informationsbeschreibung wird auch als konzeptueller Entwurf bezeichnet.

3. Festlegung der aktuellen Strukturparameter für das konkrete Datenbanksystem

Ein konkretes DBMS fußt auf einem bestimmten Datenmodell (hierarchisch, netzwerkorientiert, relational). Es bietet festgelegte Datenstrukturen an. Es verfügt über eine besondere Datendefinitionssprache, die festlegt, welche Datenstrukturen verarbeitet werden können. Aufgabe dieses Schrittes ist es, die im vorangegangenen formalisiert dargestellten Informationsbeschreibungen in die aktuellen Strukturparameter des DBMS zu übertragen. Da die physische Speicherung noch keine Rolle spielt, nennt man es logisches Datenmodell.

4. Festlegung von Einzelheiten der physischen Realisierung

Welche Speichereinheiten sollen im Betrieb des Datenbanksystems verwendet werden. Festlegung der Zugriffspfade, des Speicherplatzes je Informationseinheit (Attribut), Suchkriterien.

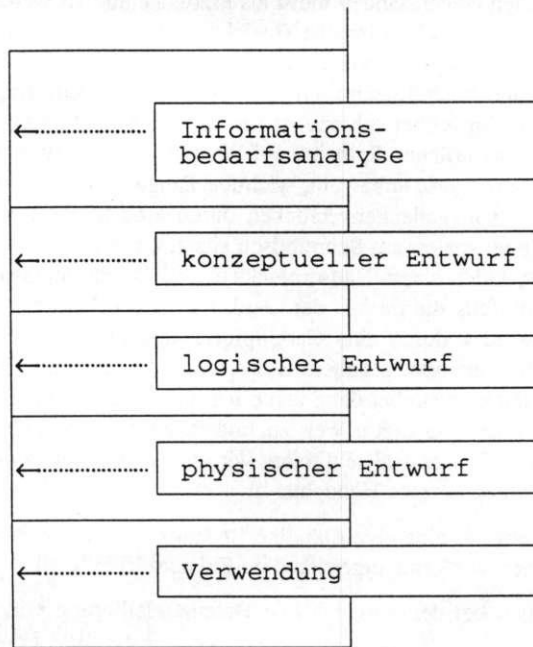
5. Entwurfsänderung.

Treten in irgendeiner Phase Probleme auf, die Korrekturen einer vorangegangenen Phase erfordern, so wird dorthin zurückgegangen und von dort erneut sequentiell Punkt für Punkt abgearbeitet. Das wird in Abb. 2 durch die gepunkteten Pfeile symbolisiert.

Konzeptueller Entwurf

Die folgenden Ausführungen beziehen sich auf die Phase des konzeptuellen Entwurfs, der eigentlich schöpferischen Phase. Hier wird die Qualität der entstehenden Software entscheidend beeinflusst. Leider wird ihr in der Praxis oft nicht die notwendige Aufmerksamkeit gewidmet, was, wie aus dem bisherigen

Abb. 2 Phasenkonzept [5]



Ausführungen hervorgeht, zu erhöhtem Aufwand bei Änderungen und zu ungünstigen Datenstrukturen führen kann. Es ist ein semantisches Datenmodell (SDM) zu schaffen, das uns helfen soll, die abzubildende Miniwelt besser zu durchdringen und präziser zu beschreiben. Der Begriff Modell wird in sehr unterschiedlicher Weise benutzt. Ein Modell im betrachteten Kontext ist die vereinfachte Nachbildung eines originalen (realen oder hypothetischen) Untersuchungsobjektes, die im Hinblick auf ein gegebenes Untersuchungsziel alle relevanten Eigenschaften des Originals aufweist.

In der Literatur sind eine Vielzahl von SDM beschrieben worden, die sehr häufig auf dem allgemeinen Systembegriff der Kybernetik beruhen. Ein System wird als eine Menge von Objekten unserer Anschauung oder unseres Denkens zwischen denen wohldefinierte Beziehungen bestehen, gesehen. Bekannteste Klasse der SDM ist das Entity Relationship Model (ERM). Im folgenden wird die Variante beschrieben, die auf Wong/Katz zurückzuführen ist.

Entscheidende Begriffe sind *Gegenstand*, auch *Entity* genannt, und *Beziehung*. Der Begriff *Gegenstand* dient der Bezeichnung von Objekten unserer Anschauung und unseres Denkens. Bei einer ersten, verbalen Beschreibung einer Miniwelt treten deren Gegenstände als Subjekte und Objekte von Aus-

sagesätzen auf. Eine *Beziehung* stellt einen Zusammenhang zwischen zwei oder mehreren *Gegenständen* dar. Bei einer verbalen Beschreibung treten Beziehungen zwischen Gegenständen meist als Prädikate auf. Es werden nicht die einzelnen Gegenstände oder einzelne Beziehungen beachtet, sondern jeweils Typen (Klassen) von Gegenständen, bzw. Beziehungen.

Die Begriffe sind noch zu schwach und unscharf und bedürfen einer Konkretisierung. Jede Umweltbetrachtung ist bereits eine Modellierung. Es werden Begriffe (z.B. der natürlichen Sprache) auf Wahrnehmungen angewandt. Dabei sind wir es gewohnt, ganz unbewußt, wichtige Dinge von unwichtigen zu abstrahieren. Wenn man gedankenversunken durch eine Straße einer Großstadt geht und sich wenig später, am Schreibtisch sitzend, versucht zu erinnern, was man gesehen hat, fallen einem Häuserblocks ein, vielleicht ungünstig parkende Autos, aber keinesfalls die Farben der Gardinen an den Fenstern oder ob das unbebaute Grundstück durch eine Gartenpforte verschlossen war oder nicht. Wenn ich jedoch, ohne über genauere Ortskenntnis zu verfügen, einer der umliegenden Gaststätten zustrebe, dann achte ich auf Durchgänge, Gartenpforten und Zäune, um einen günstigen Weg zu finden, an die ich mich später auch erinnern kann. Im Unterbewußtsein teilen wir unsere Umgebung ein, abhängig von Ziel oder Zweck unseres Handelns, in

- Objekte, denen wir eine eigenständige Bedeutung zuerkennen und
- Objekte, denen wir keine eigenständige Existenz zubilligen.

So geht man auch bei der semantischen Datenmodellierung vor. Man unterscheidet:

- unabhängige (starke) Gegenstände (strong entities), das sind Gegenstände mit eigenständig interessierender Existenz und
- abhängige (schwache) Gegenstände (weak entities), Gegenstände, deren Existenz nur im Zusammenhang mit anderen Gegenständen gegeben ist.

Starke Gegenstände werden auch nur als *Gegenstände* bezeichnet, schwache Gegenstände als *Werte*, da sie i.a. nur aus einem Wert bestehen. Werte werden zu Wertemengen (-typen) zusammengefaßt. Gegenstände, die mit den gleichen Wertemengen, also mit den gleichen Typen von schwachen Gegenständen, in Beziehung stehen, werden als gleichartige Gegenstände gesehen. Gleichartige Gegenstände bilden einen Gegenstandstyp.

Kehren wir zum Beispiel zurück, so sind Häuser (starke) Gegenstände. Sie haben eine Außenfassade (Putz, Beton, Klinker,...), Hausnummern (Zahlen) und Haustüren (Farben), die alle Werte (schwache Gegenstände) darstellen, da sie nur im Zusammenhang mit dem Haus von Interesse sind. Drei verschiedene Wertemengen treten auf: Art der Fassade, Hausnummer und Farbe der Haustür. Die einzelnen Ausprägungen 'Haus' werden zum Gegenstandstyp Haus zusammengefaßt, da sie gleichenartige Beziehungen zu Wertemengen haben.

Von der Differenzierung der Gegenstände abgeleitet, kann man auch Beziehungen klassifizieren. Die Beziehung zwischen zwei (starken) Gegenständen

nennt man *Assoziation*, die zwischen einem Wert und einem Gegenstand *Attribut*. Haus und Hausnummer sind z.B. attributiv verbunden. Assoziationen werden selbst wieder als wichtig angesehen. Aus einem anderen Blickwinkel könnte man sie auch als Gegenstände sehen. Daher wird es zugelassen, daß ihnen Werte zugeordnet werden können, d.h. sie sind attributierbar. Stellen wir uns vor, einige der Häuser sind direkt durch einen Gang im Keller miteinander verbunden. Das könnte man als Assoziation zwischen zwei Vertretern des Gegenstandstyps Haus sehen. Formalisiert könnte man es so schreiben:

(Haus)	(ist verbunden mit)	(Haus)
Subjekt	Prädikat	Objekt

Die Assoziation (ist verbunden mit) könnte man attributieren, indem man eine Aussage trifft über die Art der Verbindung: 'verschließbare Tür', 'verriegelbare Tür', 'ohne Tür'.

Assoziationen können ganz verschiedener Art sein. Im Beispiel war es eine Beziehung zwischen je zwei Vertretern des gleichen Typs von Gegenständen. Es ist natürlich auch möglich Assoziationen zwischen verschiedenen Gegenstandstypen herzustellen. Nicht alle Elemente des Gegenstandstyps müssen auch tatsächlich eine Beziehung eingehen. Darauf wird an späterer Stelle noch genauer eingegangen sein.

Seien G_1 und G_2 Gegenstandstypen zwischen denen eine Assoziation besteht. Geht jedes einzelne Element von G_1 eine Assoziation ein und zwar nur genau einmal, so nennt man das eine *Bindung* von G_1 an G_2 . Um in einem späteren Schritt die Übertragung des ERM formalen Regeln folgend in Datenstrukturen vornehmen zu können, sind folgende Einschränkungen zu machen:

- Attribute für Bindungen sind nicht zugelassen! Das läßt sich damit begründen, daß ein Attribut der Beziehung auch immer ein Attribut des gebundenen Gegenstandes ist.
- Durch ein Attribut wird genau ein Wert zugeordnet.

Das hier vorgestellte Entity Relationship Model verwendet die fünf Kategorien *GEGENSTAND*, *WERT*, *ASSOZIATION*, *ATTRIBUT* und *BINDUNG*. Alle für eine gegebene Aufgabenstellung wesentlichen Erscheinungen eines Ausschnitts der realen Welt müssen auf diese Kategorien abgebildet werden. Es empfiehlt sich eine formalisierte Darstellungsweise, Tabellen und Graphiken, wie später am konkreten Beispiel gezeigt werden wird, zu benutzen. Diese Kategorien lassen sich direkt, d.h. durch Anwendung formaler Regeln, in eine Datenbankstruktur übertragen.

Entwurf eines Datenmodells für GEOHIST

Am Beispiel des Projektes GEOHIST des Zentrums für Historische Sozialforschung (ZHSF) soll, in verkürzter Form, ein semantisches Datenmodell durch Anwendung des Entity Relationship Model hergeleitet werden. Das Projekt wurde schon an anderer Stelle beschrieben. Hier nur soviel, wie für das Verständnis des Modellierungsvorganges notwendig ist.

Es ist ein Informationssystem aufzubauen, daß alle politischen Verwaltungseinheiten Deutschlands, bzw. der deutschen Territorialstaaten, flächendeckend ab Kreisebene, von 1815 bis 1939 enthält. Es soll die administrative Struktur der Einheiten und alle territorialen Veränderungen enthalten. Aufzunehmende Merkmale sind Name der Einheit, Fläche, Einwohnerzahlen, sowie zentrale ökonomische und demographische Variablen. Die administrativen Einheiten sollen graphisch in Form von geographischen Karten darstellbar sein. Das zu entwickelnde Grundmodul soll so organisiert sein, daß die verschiedensten Nutzer ihre eigenen Merkmale hinzufügen können.

Betrachten wir die Menge der administrativen Einheiten (AE). Es sind historisch gewachsene, von Menschen geschaffene Verwaltungseinheiten. Als solche haben sie zu jedem Zeitpunkt ihres Vorhandenseins einen Namen und ein eindeutig definiertes Territorium. Gleichartige Einheiten sind zu übergeordneten Verwaltungseinheiten zusammengefaßt und bilden ein hierarchisch aufgebautes System. Sie unterliegen vielfältigen Veränderungen. Durch administrative Entscheidungen können sie aufgelöst, umstrukturiert oder neu gegründet werden. Einige administrative Einheiten sind über den gesamten Betrachtungszeitraum von 1815 bis 1939 vorhanden. Andere kommen später hinzu, z.B. durch Umstrukturierung bestehender Einheiten oder durch Zugewinn nichtdeutschen Territoriums und Begründung neuer administrativer Einheiten. Einige hören auf zu existieren, weil sie mit anderen Einheiten zusammengelegt wurden oder weil sie als deutsches Territorium verloren gingen.

Es scheint sinnvoll, ihnen eine 'Lebenszeit' zuzuordnen, ein Zeitintervall, in dem sie als administrative Einheit existieren. Für aufgelöste oder gegründete Einheiten ist dieses Vorgehen plausibel. Wie ist es aber, wenn das Territorium einer beliebigen administrativen Einheit durch Zu- oder Abgang mehr oder minder verändert wird? Ab welchem Zuwachs ist sie noch als die gleiche administrative Einheit anzusehen, die sie vorher war? Wie ist einzuschätzen, wann eine administrative Einheit bei territorialer Veränderung eine andere wird?

Bei der Vereinigung zweier administrativer Einheiten kann das neue Gebilde den Namen einer der ursprünglichen Einheiten annehmen. Ist es aber noch dieselbe administrative Einheit? Sie trägt ihren Namen, kann aber aufgrund des veränderten Territoriums eine völlig neue Qualität darstellen. Diese Schwierigkeit soll durch Abb. 3 illustriert werden. Führt man eine Analyse der natürlichen Bevölkerungsentwicklung durch, so sind Schlußfolgerungen über die Ent-

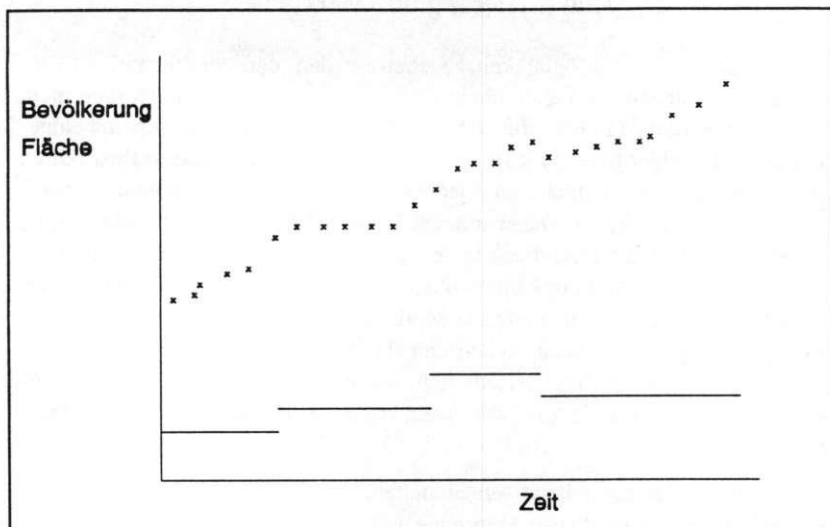


Abb. 3

Wicklung der Bevölkerung ohne aus territorialen Änderungen abgeleitete Korrekturen i.a. nicht sinnvoll. Nur innerhalb eines Zeitintervalls, in dem die Einheit ein konstantes Territorium besitzt, ist die tatsächliche natürliche Entwicklung der Bevölkerung nachzuvollziehen.

Es liegt der Gedanken nahe, nicht schlechthin administrative Einheiten zu betrachten, sondern Einheiten mit konstantem Territorium. Wenn das Territorium einer administrativen Einheit verändert wurde, sehen wir in der neuen Einheit ein anderes Objekt, auch wenn es den gleichen Namen trägt und verwaltungsrechtlich die gleiche administrative Einheit geblieben ist. In diesem Sinne gibt es z.B. nicht den Regierungsbezirk Merseburg als administrative Einheit, sondern den Regierungsbezirk Merseburg von t_0 bis t_1 oder den Regierungsbezirk Merseburg von t_1 bis t_2 , wobei die t_i für Zeitpunkte stehen zu denen sich das Territorium änderte. Um nicht jedesmal explizit darauf hinweisen zu müssen, daß nicht administrative Einheiten, sondern administrative Einheiten in einem Zeitintervall mit konstantem Territorium gemeint sind, soll hier dafür der Begriff des 'administrativen Objektes' (AO) eingeführt werden. Bei der Herleitung des Entity-Relationship-Models (ERM) wird eine noch genauere Definition vorgenommen werden müssen.

Klassifizierung der Merkmale

Dem Nutzer muß die Möglichkeit gegeben werden, den administrativen Einheiten seine eigenen Merkmale hinzufügen zu können. Es kann natürlich nicht zugelassen werden, daß über die DDL des DBMS in das interne Schema eingegriffen wird, da hier bereits kleine Fehler zu vollständigem Datenverlust führen können. In der Entwurfsphase sind jedoch die konkreten Merkmale der einzelnen Nutzer nicht bekannt. Daher müssen Wege zur Definition und Einbindung von Merkmalen in die Datenbank in der Nutzungsphase vorgesehen werden. In der Phase der Modellierung können höchstens Merkmalstypen angegeben werden, die stellvertretend für konkrete Merkmale stehen. Der Nutzer muß dann aus einer Menge von Standardfällen den für ihn geeignetsten Weg des Hinzufügens bestimmen. Dazu ist erforderlich, Klassen von Merkmalen herauszuarbeiten, die typisch für die zu erwartenden realen Merkmale sind. Man betrachtet 3 Mengen:

- die Menge der administrativen Einheiten,
- die Menge der möglichen Merkmale,
- die Menge der Zeitpunkte.

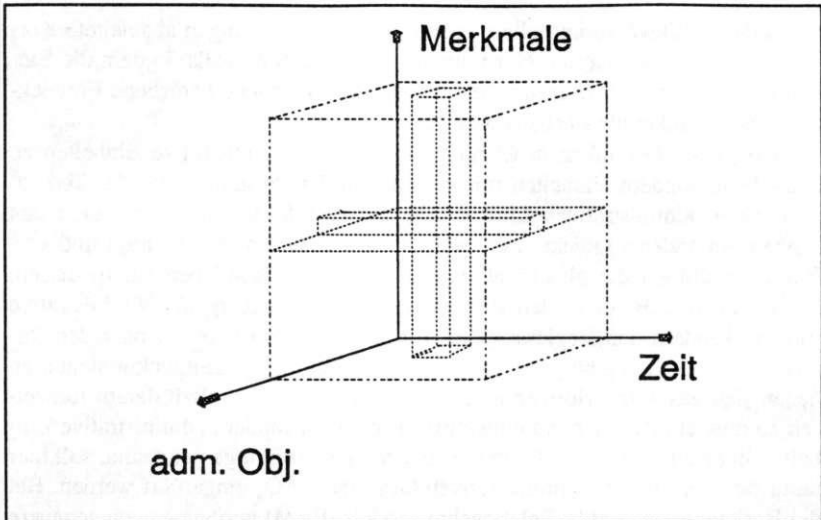


Abb. 4

Die erste Menge ist endlich, eindeutig determiniert und soll in der Datenbank vollständig erfaßt werden. Die zweite ist eine zumindest theoretisch unendliche, diskrete Menge. Die Zeit ist zwar ein Kontinuum, die Merkmale werden

jedoch zum (Tages-) Datum erfaßt. Daher ergibt sich für die Zeit ebenfalls eine endliche Menge. Das Kreuzprodukt dieser drei Mengen ist die Menge der administrativen Einheiten, für die zu bestimmten Zeitpunkten Werte für gewisse Merkmale vorliegen. In Abb. 4 ist dieser Zusammenhang graphisch dargestellt.

Eine Speicherung des Sachverhaltes als dreidimensionale Matrix scheidet wegen der schwachen Besetzung der Matrix aus: Nicht alle administrativen Einheiten gehen über die volle Zeitspanne, nicht alle Merkmale liegen für alle Zeitpunkte und administrativen Einheiten vor.

Es gibt Merkmale, für die es typisch ist, daß zu gewissen Zeitpunkten für viele administrative Einheiten Realisierungen vorliegen, z.B. die durch Zählung zu einem Stichtag erhobenen Bevölkerungsdaten. Diese Merkmale (Zensusdaten) treten i.a. in Gruppen auf, da sie in einer gemeinsamen Zählung zu einem fixierten Zeitpunkt erhoben wurden. Charakteristisch ist hier

ein Zeitpunkt,
viele Einheiten,
einige wenige Merkmale.

Andere Merkmale sind nur für einige wenige administrative Einheiten relevant, liegen aber zu relativ vielen Zeitpunkten vor (Zeitreihen). Typisch ist die folgende Situation:

viele Zeitpunkte,
wenige Einheiten,
ein (oder mehrere) Merkmal(e).

Die drei Grundmengen administrative Einheiten, Zeit, Merkmale lassen sich in vielfältiger Weise kombinieren, so daß, wie in Abb. 4 dargestellt, vertikale oder horizontale Balken, Säulen oder Ebenen entstehen. Die in der Praxis häufigsten und wichtigsten sind jedoch Zeitreihen und Zensusdaten (Querschnitte). Ein Querschnitt entspricht der Darstellung im ersten Fall (ein Zeitpunkt, viele AO, mehrere Merkmale). Eine Zeitreihe läßt sich in Abb. 4 als horizontaler Balken interpretieren. Alle anderen Zusammenstellungen der Grundmengen werden auf diese Typen zurückgeführt

Grundsätzlich läßt sich jeder Punkt im Raum der drei Grundmengen durch beide Betrachtungsweisen darstellen: als Querschnitt der nur für eine Einheit erhoben wurde oder als Zeitreihe, die nur einen Zeitpunkt enthält. Wenn jeder einzelne Punkt erfaßt werden kann, ist auch jede beliebige Ausdehnung eines Merkmals darstellbar. Die Effektivität des Systems hängt davon ab, in welchem Maße Querschnitt und Zeitreihe typisch für den tatsächlich auftretenden Datenbestand sind und ob der Nutzer seine Merkmale auch optimal den Typen zuordnet.

Entity-Relationship-Model

Auf den ersten Blick scheint die administrative Einheit ein hervorgehobener, für die Betrachtung wichtigster Gegenstandstyp zu sein. Wird sie jedoch über den gesamten Zeitraum als eine Entity betrachtet, treten erhebliche Schwierigkeiten auf. Zum einen kann durch territoriale Zu- und Abgänge die Vergleichbarkeit in der zeitlichen Entwicklung verloren gehen. Zum anderen können administrative Einheiten ihre Zugehörigkeit zu übergeordneten Einheiten verändern, z.B. kann ein Kreis einem anderen Regierungsbezirk unterstellt werden. Dieser neuen Zuordnung einer administrativen Einheit zu einer andern übergeordneten administrativen Einheit wird bei unserer Betrachtung eine wesentliche Bedeutung beigemessen. Geht beispielsweise ein Kreis von einem wirtschaftlich unterentwickelten Regierungsbezirk zu einem wirtschaftlich entwickelten über, so kann das für einige seiner Merkmale von gravierender Bedeutung sein.

Die Gliederung der administrativen Einheiten ist eine Baumstruktur². Diese Baumstruktur ist für jeden Zeitpunkt eindeutig gegeben. Sie ist aber in der zeitlichen Entwicklung veränderlich. Der Baum unterliegt einer Vielzahl von Veränderungen, aber nicht notwendig solchen, wie es dem natürlichen Wachsen eines Baumes entsprechen würde. Änderungen der Struktur führen nicht nur dazu, daß an den Enden neue Knoten angefügt werden, sondern es treten auch im Innern des Baumes Veränderungen auf.

Bevor festgelegt wird, was als Entity zu betrachten ist, sollen die Veränderungen, denen eine administrative Einheit unterliegt, dargestellt werden:

1. Zu- oder Abgang von Territorium von/nach anderen administrativen Einheiten,
2. Auflösung der administrativen Einheit, d.h. das Territorium wird einer oder mehreren anderen administrativen Einheiten zugeordnet,
3. Entstehen einer neuen administrativen Einheit,
4. Zuordnung zu einer anderen übergeordneten Einheit,
5. Veränderung der Struktur innerhalb einer Einheit.

Nicht nur das Territorium, auch andere Merkmale können sich verändern. Eine starke Zunahme der Wohnbevölkerung durch natürliches Wachstum oder durch Zuwanderung kann aus einer bestimmten Sicht eine qualitative Veränderung der administrativen Einheit bedeuten. Ob solchen Veränderungen eine entsprechende Bedeutung beigemessen wird, hängt von der Aufgabenstellung und der damit verbundenen Sicht auf die abzubildenden realen Welt ab.

² Unter einem Baum versteht man eine Menge von Knoten, die durch eine Menge von gerichteten Kanten verbunden sind. Der Knoten, von dem die Kante ausgeht, wird Vorgänger, der andere Nachfolger genannt. Es gelten folgende Bedingungen: Es existiert genau ein ausgewählter Knoten der keinen Vorgänger besitzt (Wurzel). Alle anderen Knoten besitzen genau einen Vorgänger.

Unter dem Gegenstand administratives Objekt (AO) soll eine administrative Einheit verstanden werden, für die, betrachtet in einem Zeitintervall, die folgenden Bedingungen erfüllt sind:

- Für die betrachtete Einheit treten im Innern des Zeitintervalls keine Änderungen nach Punkt 1 bis 4 (oben) auf.
- Für die Einheit im unmittelbar übergeordnetem Baum der administrativen Struktur, d.h. auf dem Pfad zur Wurzel, treten höchstens Veränderungen nach Punkt 1 oder 5 auf, von denen jedoch die betrachtete Einheit nicht direkt betroffen ist
- Zu den Zeitintervallgrenzen treten Ereignisse ein, wie sie in der Bedingung 1 und 2 ausgeschlossen wurden.

Die erste Forderung ist offensichtlich. Es ist nur sinnvoll, die Entwicklung bestimmter Merkmale eines Territoriums bei konstanter Fläche zu betrachten. Die zweite besagt daß es für eine administrative Einheit eine qualitative Veränderung bedeuten kann, wenn sich die Unterstellung zu ihren übergeordneten Einheiten, d.h. ihre sie umfassenden administrativen Einheiten, ändert. Sie muß nach einer solchen Änderung als ein anderes AO gesehen werden.

Die sechsstufige Struktur³ - Ort, Amt Kreis, Regierungsbezirk, Provinz, Land - ermöglicht eine Klassifizierung der AO. Jede dieser Klassen könnte als ein Entity-Typ betrachtet werden. Zwei Gründe sprechen dafür, es nicht zu tun. Für alle liegt eine ähnliche Menge Attribute vor, so daß es naheliegend ist zu generalisieren. Der zweite Grund ist, daß die Grenzen zwischen den Klassen nicht starr sind. Z.B. können aus Ländern Provinzen oder Regierungsbezirke werden. Das könnte, wie aus der Beschreibung der Beziehungen im folgenden Absatz ersichtlich wird, erhebliche Probleme nach sich ziehen. Man kann 3 Arten von Beziehungen zwischen AO unterscheiden:

- *ADMINISTRATIVE ZUGEHÖRIGKEIT*. Das sind solche Beziehungen zwischen AO, die sich in der Baumstruktur widerspiegeln (s. Abb. 5). Zu jedem Objekt eines bestimmten administrativen Niveaus gehören ihm untergeordnete Objekte. Jedes Objekt (außer dem Niveau 'Land') gehört zu einem übergeordneten Objekt. Z.B. gehört ein Ort zu einem Kreis, ein Kreis zu einem Regierungsbezirk usw.. Sie bestehen per Definition während des dem AO zugeordneten Zeitintervalls unverändert.
- *VERÄNDERUNG VON TERRITORIUM*. Durch eine Flächenänderung endet ein Objekt bzw. ein neues beginnt. Eine administrative Einheit verliert Territorium, eine andere bekommt es hinzu. Dadurch wird eine Beziehung zwischen den beiden Einheiten realisiert (s. Abb. 6). Aus *beiden* Einheiten entstehen neue AO (Änderung nach Punkt 1).

³ Es wird hier die administrative Struktur Preußens auf das Gesamtterritorium übertragen. Das bereitet keine Probleme, wenn die entsprechende administrative Struktur dort, wo sie nicht in dieser Tiefe auftrat, unberücksichtigt bleibt.

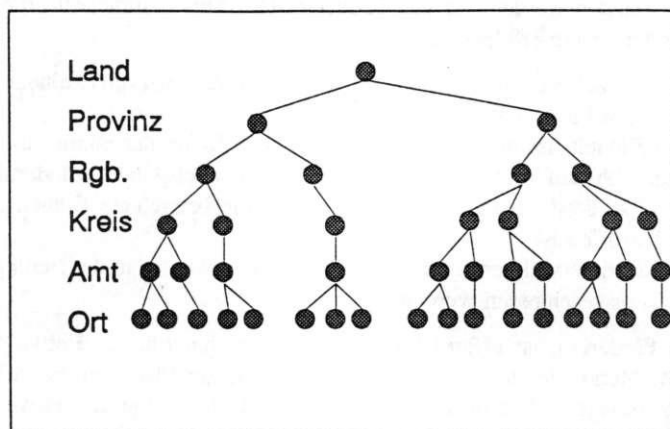


Abb. 5

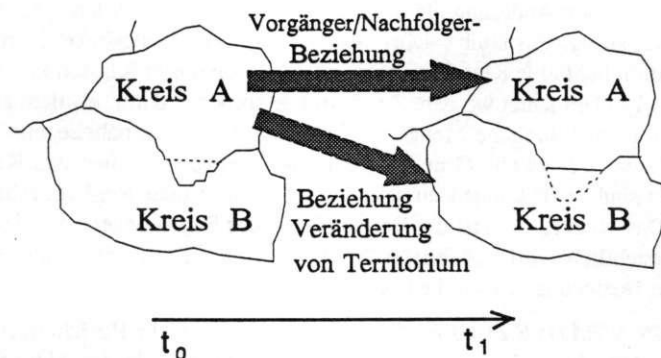


Abb. 6

- **VORGÄNGER/NACHFOLGER.** Wenn eine Einheit zum Zeitpunkt t_0 Territorium hinzu bekommt, hört sie auf als administratives Objekt zu existieren. Mit dem Zeitpunkt t_1 entsteht aus ihr ein neues administratives Objekt (s. Abb. 6). Dieses neue Objekt kann unter bestimmten Voraussetzungen als Nachfolger des alten betrachtet werden.

Die Sinnförligkeit, ob es tatsächlich der Nachfolger ist, ist ein administratives Problem, keine formale Modellierungsfrage. Ein Regierungsbezirk X beispielsweise, der zum Zeitpunkt t_0 durch nicht näher interessierende Maßnahmen zu einem Zeitpunkt t_1 territorialen Zuwachs bekommt, wird ein anderes Objekt. Zwischen X_0 und X_1 besteht natürlich ein enger Zusammenhang. In einem

anderen Konsens würden wir X_0 und X_1 als dieselbe Einheit ansehen, es ist der gleiche Regierungsbezirk geblieben. Genauer gesehen ist jedoch X_0 aus X_1 und einem gewissen Teil eines anderen Objektes, nennen wir den territorialen Zugewinn δX , hervorgegangen. Formal könnte man es so darstellen:

$$X_1 = X_0 + \delta X$$

X_1 wird als Nachfolger von X_0 aufgefaßt, wenn er das im administrativen Sinne ist. Ein Nachfolger muß jedoch nicht existieren. Wird eine Einheit geteilt, so kann jedes territoriale Teilstück als nur Zugang zu einer anderen Einheit gesehen werden, also als Beziehung nach Punkt zwei.

Alle territorialen und administrativen Änderungen lassen sich auf die drei genannten Grundbeziehungen zurückführen, gegebenenfalls durch Zerlegung eines komplexeren Vorgangs in mehrere Einzelbeziehungen. Die Beziehung VERÄNDERUNG VON TERRITORIUM und die VORGÄNGER/NACHFOLGER-Beziehung werden zu einem bestimmten Zeitpunkt wirksam. Die erste Beziehung korrespondiert mit einer Flächenveränderung der Größe δX . Diese Werte stellen Attribute dar. Für die Beziehung ADMINISTRATIVE ZUGEHÖRIGKEIT gilt, daß jedes AO zu jedem Zeitpunkt genau ein übergeordnetes AO besitzt. Solche funktionalen Abbildungen hatten wir Bindungen genannt. Jedes Entity AO steht mit zwei Zeitpunkten in Beziehung, dem Startpunkt und dem Endpunkt. Diese Zeitpunkte sind im Sinne des ERM schwache Gegenstände (Werte). Mit Hilfe der genannten Beziehungen läßt sich der Gegenstand 'administratives Objekt' auch wie folgt beschreiben: Für einen Zeitraum (t_0, t_1) sind folgende Bedingungen erfüllt:

- zum Zeitpunkt t_0 und t_1 treten Beziehungen der zweiten Art auf (Wechsel von Territorium) oder im übergeordneten Baum wird eine Beziehung der ersten Art (administrative Unterstellung) wirksam;
- das für t_0 und t_1 gesagte trifft für keinen Zeitpunkt im Innern des Intervall (t_0, t_1) zu.

Man sieht an dieser Definition, daß jedem AO ein Existenzintervall zugeordnet ist. Administrative Objekte haben einen Bezeichner, den geographischen NAMEN. Er ist für jedes Objekt eindeutig gegeben, allerdings können mehrere Objekte den gleichen Namen tragen. 'Hannover' z.B. bezeichnet einen Ort, einen Kreis einen Regierungsbezirk, eine Provinz und ein Land. Dieser Gegenstand 'Name' besteht aus einer einzigen Zeichenfolge, einem einzelnen Wert, der nur im Zusammenhang mit AO von Bedeutung ist. Deshalb ist der Name ein schwacher Gegenstand. Umgangssprachlich werden AO benannt durch

- den Namen des Territoriums,
- das administrative Niveau (Ort, Amt/Gerichtsbezirk, **Kreis**, **Regierungsbezirk**, Provinz, Land)
- ein Zeitintervall.

Leider eignet sich dieses Tripel nicht als Identifikator, u.a wegen der oben beschriebenen Eigenschaften des Namens. Daher müssen künstliche oder Surrogat-Namen, die den Nachteil des geographischen Ortsnamens überwinden, eingeführt werden. In der EDV werden solche Bezeichner SCHLÜSSEL oder 'key' genannt. Hier soll ein Schlüssel benutzt werden, der die *administrative Einheit* identifiziert, nicht das administrative Objekt. Gemeinsam mit dem Zeitintervall, genauer mit dem Startpunkt, bildet der SCHLÜSSEL einen Identifikator für administrative Objekte. Eine Kombination von Attributen, die ein Objekt eindeutig identifizieren, nennt man Primärschlüssel.

Eine Gewerbe-, Volks- oder Berufszählung stellt eine Messung gewisser Merkmale nach bestimmten Regeln zu einem Zeitpunkt dar. Sie kann für alle zu diesem Zeitpunkt gültigen AO erfolgt sein oder nur für einige vorliegen. Es ist sowohl möglich, einen Meßwert (z.B. Bevölkerung) zu erheben, als auch Tupel von Meßwerten (z.B. Bevölkerung, Fläche, Gewerbetätige, Berufstätige in der Industrie, Berufstätige in der Landwirtschaft). Das hängt von den konkreten Zensusdaten ab. Hier soll der Aspekt im Vordergrund stehen, daß die verschiedenen Merkmale eines Objektes als ein Meßpunkt betrachtet werden. Sieht man MESSPUNKTE als Entities, so erhält man die Assoziation 'QUERSCHNITT i' zwischen AO und diesen Meßpunkten. Der Zeitpunkt stellt eine attributive Verbindung zu dieser Assoziation dar.

Zählungen können nach ganz unterschiedlicher Methodik zu verschiedenen Zeitpunkten und für unterschiedliche Merkmale durchgeführt worden sein. Es gibt also nicht schlechthin einen Beziehungstyp 'Querschnitt', sondern nach bestimmten Regeln zu einem bestimmten Zeitpunkt durchgeführte Zählungen. Wir haben damit nicht einen, sondern eine Reihe von Beziehungstypen zu betrachten:

'Querschnitt 1', 'Querschnitt 2',...usw..

Die Werte der Merkmale für die der Zensus erhoben wurde, sind attributiv verbunden mit den Meßpunkten.

Etwas anders verhält es sich mit ZEITREIHEN. Eine Zeitreihe besteht aus einer Menge von Paaren (T_i, X_i), wobei T_i Zeitpunkte und X_i Merkmalsrealisierungen bedeuten. Für unsere Betrachtung soll eine Zeitreihe als ganzes gesehen werden, jede einzelne Zeitreihe als Gegenstand des Entitytyps Zeitreihe. Es wäre möglich, Zeitreihen jeweils auf nur ein AO zu beziehen oder aber mehrere administrative Objekte innerhalb einer Zeitreihe zuzulassen. Es würde der Vorgänger/Nachfolger-Beziehung Rechnung tragen, eine Zeitreihe für mehrere Objekte zusammenzustellen. Hier soll jedoch ein Merkmal für eine administrative Einheit in zeitlicher Entwicklung verfolgt werden, ohne Ereignisse berücksichtigen zu müssen, die die Entstehung eines neuen AO induzieren (territoriale Veränderungen, veränderte administrative Zuordnungen). D.h. es werden nur solche Zeitreihen dem Entitytyp »Zeitreihe« zugeordnet, die sich auf ein AO beziehen. Will man die Entwicklung einer administrativen Einheit be-

trachten, sind außerhalb des Datenbestandes mehrere Zeitreihen zu verbinden. Eine Zeitreihe steht mit den beiden schwachen Gegenständen Zeitpunkt und Wert des Merkmals in attributiver Beziehung. Die Zuordnung einer Zeitreihe zu einem AO stellt eine Assoziation dar, da beide Gegenstandstypen sind.

Informationen über Dateien und Datenfelder werden vom DBMS verwaltet. Vor dem Aufbau einer Datenbank werden diese Informationen im Dialog oder mittels der DDL dem System mitgeteilt und in sogenannte Data Definition Tables abgelegt. Da in unserem Fall die Nutzer nicht in die physischen Dateistrukturen eingreifen sollen, muß auf der Entwurfsebene eine Art Katalog als Schnittstelle zwischen Nutzer und System vorgesehen werden. Diese Aufgabe erfüllt das Entity MERKMALSINFORMATION mit den Attributen Merkmalsname, Merkmalstyp und Zeitreihen-ZQuerschnittsname. Für den Typ Querschnitt wird noch der Zeitpunkt des Querschnitt hinzugezogen und für den Typ Zeitreihe der Identifikator des AO.

Diese Betrachtungsweise faßt »Merkmalsinformation« als einen komplexen Gegenstand. Dabei werden assoziative Beziehungen zwischen Merkmal, Zeitreihe und AO, sowie Querschnitt und Zeitpunkt impliziert. Man könnte auch den Gegenstand 'Merkmalsinformation' enger fassen und die assoziativen Beziehungen explizit darstellen. Da der Begriff des Merkmalkataloges eine klare Vorstellung von seiner Funktion vermittelt und man nach der Übertragung in relationale Datenstrukturen das gleiche Resultat erhält, soll hier darauf verzichtet werden.

Jede administrative Einheit hat einen eindeutigen Grenzverlauf, der sich innerhalb eines Intervalls nicht ändert. Daher kann jedem AO eindeutig eine UMRISSKARTE zugeordnet werden. Die Karten eines Aggregationsniveaus (Kreis, Rgb., Prov., Land) von AO überschneiden sich nicht und die Karten ab Niveau Kreis überdecken jeweils die übergeordneten vollständig. Man kann sich die Karten als Teile eines Puzzlespiels vorstellen. Zu bestimmten Zeitpunkten treten Veränderungen der Form der Karten auf. Es müssen einzelne, jeweils zusammenhängende Kärtchen herausgenommen und durch neue ersetzt werden, die nach außen die gleiche Form haben müssen, um sich anstelle der alten einsetzen zu lassen. Gegeneinander haben sie jedoch eine veränderte Gestalt.

Der Entitytyp »Umrisskarte« kann ohne Kenntnis der spezifischen Graphiksoftware noch nicht näher beschrieben werden. Man kann ihnen, analog zum Vorgehen bei administrativen Objekten, ein Zeitintervall zuordnen. Statt der sich aneinanderreihenden, bzw. überlagernden Karten ließe es sich vorstellen, eine Folge von Gesamtkarten als graphische Oberfläche aufzubauen. Diese Betrachtungsweise würde der Tatsache gerecht werden, daß es eine Vielzahl von Programmen gibt, die geographische Karten erzeugen, die sich zwar zoomen lassen, aber nicht Teilkarten zu einer Gesamtkarte zusammenfügen können. Allein im Regierungsbezirk Hannover treten im Zeitraum 1871 - 1939 etwa 140 territoriale Veränderungen auf. Rechnet man diese Zahl auf das ge-

samte deutsche Territorium im Zeitraum von 1815 bis 1939 hoch, müßte man wahrscheinlich mit mehr als eintausend Gesamtkarten rechnen. Der daraus resultierende Speicherbedarf wäre unvertretbar groß. Die folgende, nicht ganz vollständige Übersicht (vgl. Abb. 7) stellt Entities und Relationships dar.

Gegenstände und Beziehungen lassen sich i.a. verbal durch einfache Sätze unter Verwendung von 'Subjekt - Prädikat - Objekt' beschreiben. Das soll im folgenden für alle Beziehungen vorgenommen werden. Jede Beziehung, an der mehr als zwei Gegenstände beteiligt sind, wird in mehrere duale Beziehungen zerlegt

- ein AO ist entstanden zu einem STARTZEITPUNKT t_i
- ein AO endete zu einem ENDZEITPUNKT t_i
- einem AO ist zugeordnet eine FLAECHEN
- ein AO gehört zu einem übergeordneten AO
- ein AO hat ein nachfolgendes/vorausgehendes AO
- ein AO gewinnt/verliert Fläche δF zum Zeitpunkt t_i
- ein AO gewinnt/verliert Fläche von/an einem anderen AO
- ein AO hat einen NAMEN
- ein AO gehört einer KLASSE an (Ort, Amt Kreis, Rgb., Prov., Land)
- ein AO hat eine bestimmte Form (UMRISSKARTE)
- ein AO kann versehen werden mit einer BEMERKUNG
- Zensus t_i wird durchgeführt für ein AO
- Zensus t_i liefert eine Realisierung für Merkmale
- ein Zensus wird zu einem Zeitpunkt erhoben
- eine Zeitreihe liegt für ein AO vor
- ein Zeitreihenelement der Reihe Z_j enthält eine Zeitangabe
- ein Zeitreihenelement der Reihe Z_j enthält eine Merkmalsrealisierung
- eine Merkmalsinformation enthält eine Merkmalsbezeichnung
- eine Merkmalsinformation enthält einen File-Namen
- eine Merkmalsinformation enthält einen Typ-Bezeichnung
- eine Merkmalsinformation enthält einen Zeitpunkt (für Querschnitte)
- eine Merkmalsinformation enthält einen AO-Identifikator

Für die Übertragung des ER-Diagramms in eine relationale oder Netzwerkdatenstruktur ist der Grad der Beziehungen von Bedeutung. Unter dem Grad versteht man, wie viele Objekte des einen Gegenstandstyps mit wievielen Objekten des anderen Gegenstandstyps verbunden sein können. Dabei ist nicht die genaue Anzahl wichtig, sie kann vorab ohnehin nicht genau angegeben werden, sondern ob es sich um *genau ein* Objekt *höchstens ein* Objekt *mehrere* Objekte oder *mindestens ein* Objekt handelt. Der Grad der Beziehungen wird als Verhältniszahl angegeben. So stehen z.B. die administrativen Objekte vom Niveau Regierungsbezirk zu den administrativen Objekten vom Niveau Kreis in einer 1:n - Beziehung. Einem Regierungsbezirk können mehrere Kreise (n) zugeordnet sein, jeder Kreis gehört aber genau zu einem Regierungsbezirk.

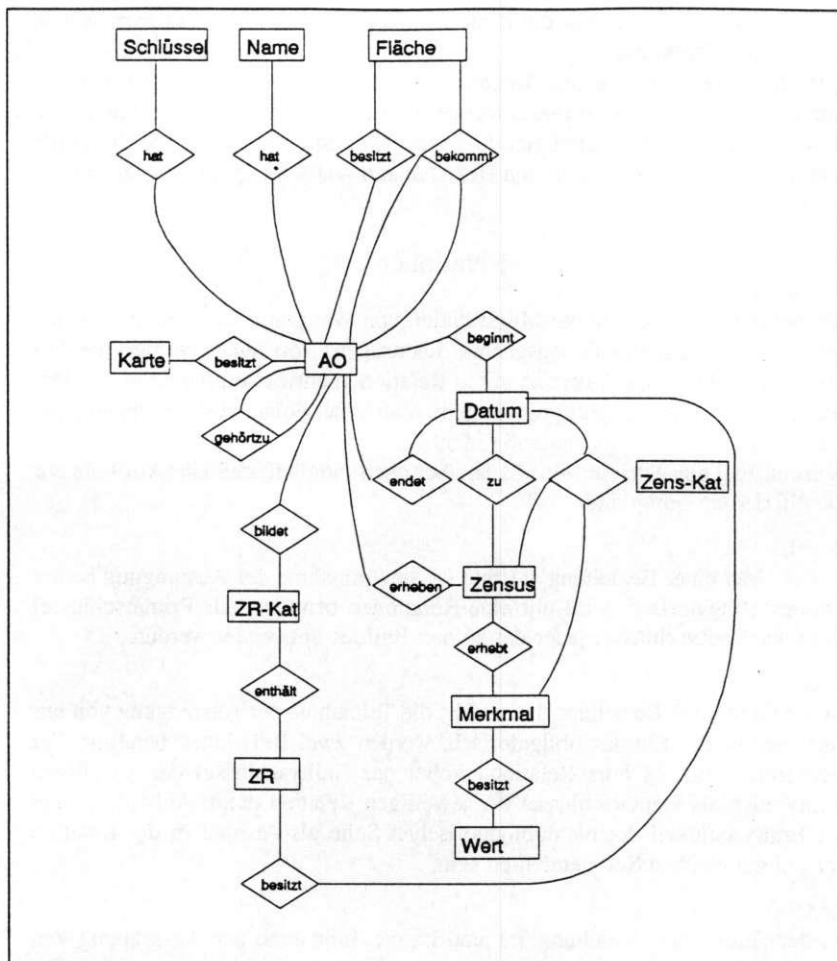


Abb. 7

In ER-Iyppdiagrammen stellt man den Grad formalisiert dar. Die Gegenstandstypen werden von Rechtecken umrahmt, die Beziehungen von Rauten, die durch Linien verbunden werden. An den Linien trägt man den Grad ab. Für genau ein oder höchstens ein Objekt wird '1' geschrieben, für mehrere trägt man die Buchstaben 'n' oder 'm' ein. Um zu symbolisieren, daß *jedes* Objekt eines Typs eine Beziehung eingeht, d.h. daß die Teilnahme jeder Ausprägung obligatorisch ist, zieht man die Linie des Rechtecks doppelt und trägt wie in Abb.6 einen Punkt ein.

In der Tabelle 1 sind alle Gegenstände und ihre Attribute aufgelistet. Es wurde nach Typ der Merkmalsinformation differenziert. Tabelle 2 enthält die Assoziationen mit den beteiligten Gegenständen und Attributen. Die durch die Merkmalsinformation realisierten Beziehungen wurden explizit angegeben.

Relationen

Wendet man das ERM in der Modifikation von Wong und Katz an, läßt es sich ohne Mühe in ein Relationenschema⁴ übertragen. Von ihnen wurden die folgenden Regeln für die Übertragung in Relationen entwickelt. Unter einem Primärschlüssel eines Objekttyps versteht man eine Folge von Attributen, die jedes Objekt dieses Typs eindeutig identifizieren. Primärschlüssel kann im einfachsten Fall ein Attribut sein. Es ist aber auch möglich, daß alle Attribute zur Identifizierung nötig sind.

Regel 1

Ist der Grad einer Beziehung 1:1 und ist die Teilnahme der Ausprägung beider Entities obligatorisch, wird nur eine Relationen benötigt. Als Primärschlüssel kann der Entityschlüssel jeder der beiden Entities verwendet werden.

Regel 2

Ist der Grad einer Beziehung 1:1 und ist die Teilnahme der Ausprägung von nur einer der beider Entities obligatorisch, werden zwei Relationen benötigt. Für jede Entity gibt es eine Relation, wobei der Entityschlüssel der jeweiligen Entity auch als Primärschlüssel der jeweiligen Relation dient. Außerdem muß der Entityschlüssel der nichtobligatorischen Seite als Attribut in der Relation der obligatorischen Seite enthalten sein.

Regel 3

Ist der Grad einer Beziehung 1:1 und ist die Teilnahme der Ausprägung von keiner der beider Entities obligatorisch, werden drei Relationen benötigt. Für jede Entity gibt es eine Relation, wobei der Entityschlüssel der jeweiligen Entity auch als Primärschlüssel der jeweiligen Relation dient, und für die Beziehung gibt es eine Relation. In dieser Relation sind unter anderem die Entityschlüssel der beiden Entities enthalten.

⁴ Relation im Sinne der Relationalen Algebra nach Codd.

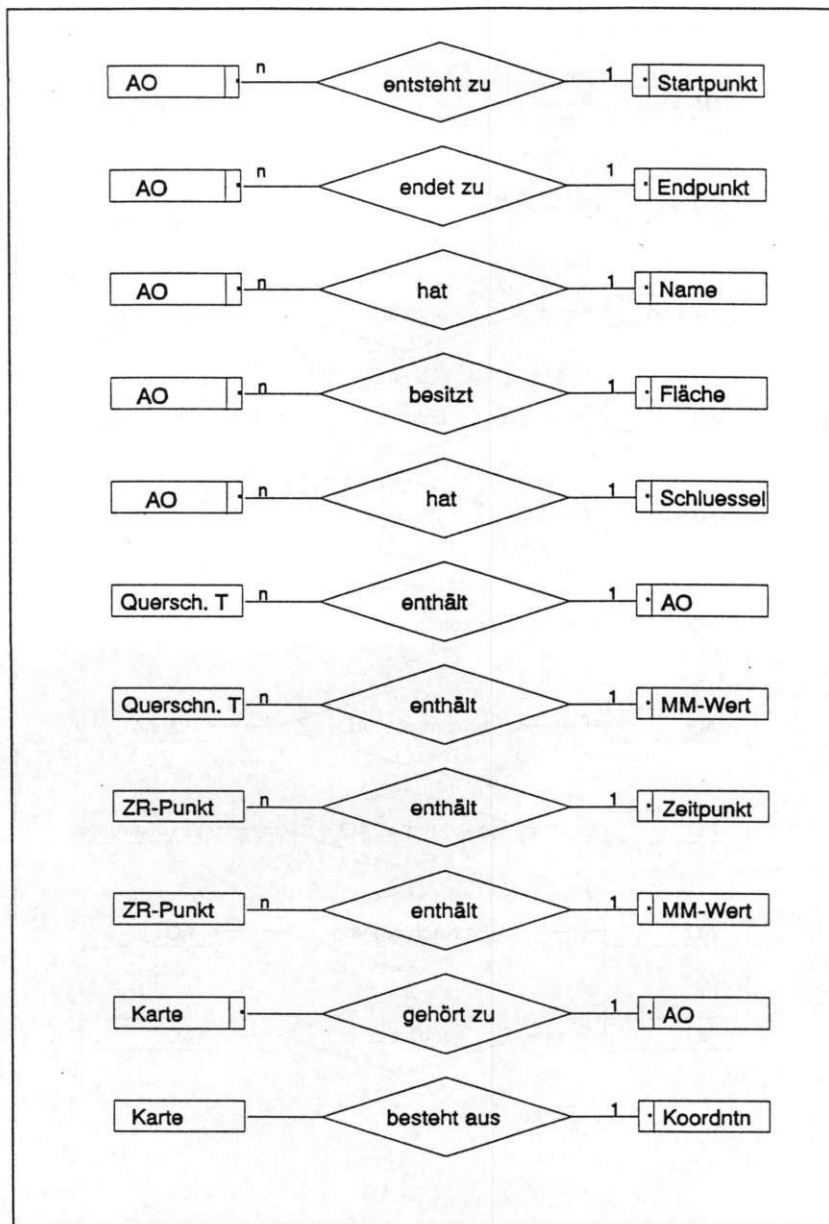


Abb. 8 ER-Diagramme

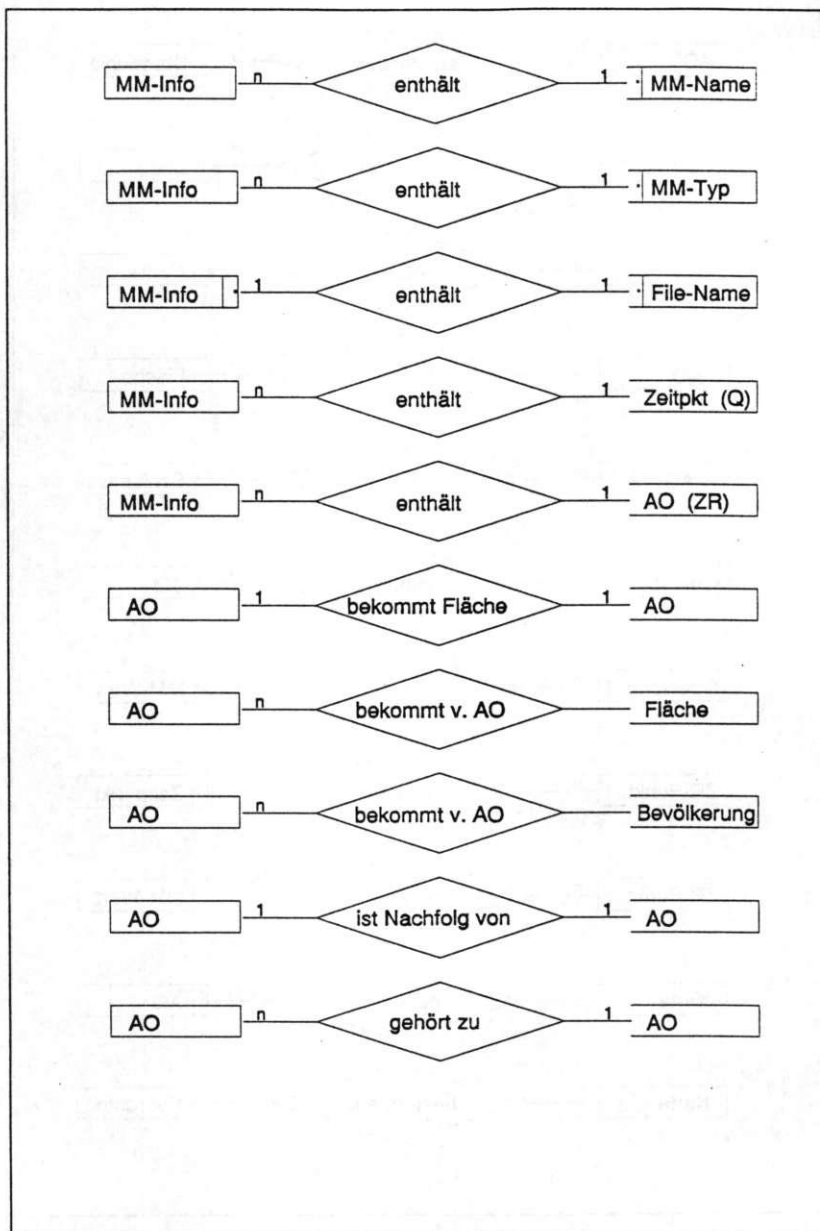


Abb. 8 (Forts.) ER-Diagramme

Tabelle 1: Entities und Attribute

Entitytyp	Attribut	Wert
AO	Schlüssel Name Klasse (Ort, Kreis, Rgb, Prov, Land) Startpunkt Endpunkt Fläche	Zeichen Zeichen Zeichen Datum Datum Zahl
Karte	Koordinaten bzgl. eines Bezugs- systems Startpunkt Endpunkt	Zahlen
Merkmais- Information (Zensus)	Merkmalsname Dateiname Stichtag der Zählung Bemerkung	Zeichen Zeichen Datum Text
Merkmais- Information (Zeitreihe)	Merkmalsname Dateiname AO-Primärschlüssel Bemerkung	Zeichen Zeichen Zeichen Text
Meßpunkt	Merkmal 1 Merkmal 2 ...	MM-Typ 1 MM-Typ 2 ...
Zeitreihe	Zeitpunkt Merkmalsrealisierung	Datum MM-Typ

Tabelle 2: Assoziationen und Entities

Assoziation	Entity	Attribut
Zugehörigkeit	AO AO	
Flächenveränderung	AO AO	Zeitpunkt Fläche
Nachfolger	AO AO	
Abbildung	AO Karte	
Querschnitt i	Meßpunkt AO	
Entwicklung eines AO	Zeitreihe AO	
Katalogeintrag Querschnittsname	Merkmalsinfo Zensus	
Katalogeintrag Zeitpunkt	Querschnitt Zeit	
Katalogeintrag Zeitreihenname	Merkmalsinfo ZR	
Katalogeintrag AO-Identifikator	Merkmalsinfo AO	

Regel 4

Ist der Grad einer Beziehung 1:n und ist die Teilnahme der Ausprägung der n-Seite obligatorisch, werden zwei Relationen benötigt. Für jede Entity gibt es eine Relation, wobei der Entityschlüssel der jeweiligen Entity auch als Primärschlüssel der jeweiligen Relation dient. Außerdem muß der Entityschlüssel der 1-Seite als Attribut in der Relation der n-Seite enthalten sein.

Regel 5

Ist der Grad einer Beziehung 1:n und ist die Teilnahme der Ausprägung der n-Seite nicht obligatorisch, werden drei Relationen benötigt. Für jede Entity gibt es eine Relation, wobei der Entityschlüssel der jeweiligen Entity auch als Primärschlüssel der jeweiligen Relation dient, und für die Beziehung gibt es eine Relation. In dieser Relation sind unter anderem die Entityschlüssel der beiden Entities enthalten.

Regel 6

Ist der Grad einer Beziehung m:n, werden drei Relationen benötigt. Für jede Entity gibt es eine Relation, wobei der Entityschlüssel der jeweiligen Entity auch als Primärschlüssel der jeweiligen Relation dient, und für die Beziehung gibt es eine Relation. In dieser Relation sind unter anderem die Entityschlüssel der beiden Entities enthalten.

Unter Berücksichtigung von Abb. 6, Tab. 1 und Tab. 2 erhält man nach einigen Zusammenfassungen folgendes relationale Schema. Die Primärschlüssel sind unterstrichen. Auf Erklärungen und Bemerkungen, die attributiv mit einzelnen Gegenständen und Assoziationen verbunden sein können, wurde verzichtet. Sie sind leicht anzufügen.

Relation AO

Schlüssel für administrative Einheiten (Surrogatname)

Startpunkt

Endpunkt

Primärschlüssel des übergeordneten AO

Klasse

Name

Fläche

Relation ZU_ABGANG

Richtung der Flächenänderung (Zu- oder Abgang)

Zeitpunkt

Primärschlüssel des ersten AO

Primärschlüssel des zweiten AO

Hache (δF)

Relation NACHFOLGER

Primärschlüssel des Vorgängers

Primärschlüssel des Nachfolgers

Relation ZENSUSKATALOG

Merkmalsname

Querschnittsname (Filename)

Stichtag

Relation QUERSCHNITTSKATALOG

Merkmalsname

Zeitreihenname (Filename)

Primärschlüssel des AO

Schlußbemerkungen

Es stellt sich natürlich die Frage, welchen Vorteil ein solches Herangehen bietet. Jeder Laie, der mit einem gebräuchlichen PC-Datenbankprogramm arbeitet, kann nach kurzem Nachschlagen im Manual seine Dateistruktur definieren und Daten eingeben. Es muß klar gesagt werden, wer eine rechteckige Datenmatrix niedriger Dimension ohne tiefergehende innere Struktur verarbeiten möchte, braucht keine aufwendige Datenmodellierung zu betreiben.

Anders verhält es sich bei komplexeren Problemen. Explikation der Begriffe, Herausstellen der wesentlichen Seiten, Formalisierung, Modellbildung sind feste Bestandteile einer akzeptierten Forschungsstrategie. Zweifellos hilft das ERM, Zusammenhänge der zu untersuchenden Welt besser zu erkennen sowie Merkmale und Erscheinungen komplexer zu erfassen. Man wird schnell erkennen, welche Daten man unter einer bestimmten Aufgabenstellung aufnehmen muß und welche nicht benötigt werden.

Aufgabenstellungen sind nicht starr. Nach eingehender Beschäftigung mit einem Forschungsgegenstand kann sich die Sicht auf die untersuchten Erscheinungen und Prozesse ändern. Solche veränderten Sichtweisen sind wesentlich einfacher und effektiver auf den Datenbestand zu übertragen, wenn man zuerst das ERM anpaßt und daraus veränderte Datenstrukturen ableitet.

Ein weiterer Vorteil ist, daß bei konsequenter Einhaltung aller Regeln die Dateien bereits normalisiert (3. NF) sind. Auf das Normalisieren soll hier nicht näher eingegangen werden. Es sei nur so viel gesagt, daß normalisierte Dateien eine für die Bearbeitung günstige Form darstellen.

Die Komplexität und der Aufwand des Projektes 'GEOHIST' zwangen uns, methodisch begründete Datenstrukturen herzuleiten. Das ERM liefert nicht *die* Datenstruktur, sondern *eine* mögliche Struktur. Bestimmt gibt es andere, vielleicht auch bessere. Dieser Vorschlag gilt als Diskussionsgrundlage, um gegebenenfalls bessere Strukturen zu finden.

Literatur:

1. Data Base Task Group of CODASYL Programming Language Committee, Report, ACM Conf. on Data Syst. Language, New York, 1971.
2. Codd, E. F.: A relational model of data for large shared data banks, Comm. ACM, 1970.
3. Codd, E. F.: Extending the database relational model to capture more meaning, ACM Trans. Database Syst, 1979.
4. ANSI/X3/SPARC Study group on data base management systems. Interim Report, FDT, ACM SIGMOD 7,2 1975.
5. Lockemann, P.C.; Schmidt, J.W. (Hrsg.): Datenbank-Handbuch. Berlin u.a., 1987.
6. Wong, E.; Katz, R. H.: Logical design and **schema conversion for relational** and DBTG databases. In: Chen, P. P. (Hrsg.): Entity-relationship approach to systems analysis and design. Amsterdam, 1980.